

AMENDMENTS to the DRAWINGS

No amendments or changes to the Drawings are proposed.

REMARKS

Reconsideration by the Examiner

We appreciate the reconsideration and rejections under 35 U.S.C. §103(a) over Burdick in view of Wocke.

Unclear Finality of Present Rejections

The first paragraph of the current Office Action indicates that our previous arguments were persuasive, and that the current Office Action is non-final. This is consistent with the fact that the previous rejections were withdrawn without our having amended the claims (we filed remarks only).

However, the summary of the Office Action form shows the action to be Final, the last several pages of the Office Action (Response to Applicant's Arguments section) states several times that the Examiner was not persuaded by the arguments, and PAIR indicates the Office Action is final.

We have attempted to contact the Examiner on two occasions to clarify the status, but have not been successful. We are presuming the rejections are final, but must also presume that our arguments (without amendment) were at least somewhat persuasive which led to the change in the reasons for rejections.

Nature of Amendment

In the present amendment, we have amended our claims directed to method embodiments of our invention, and we have cancelled all other pending claims from further consideration in this application. We are not conceding that the subject matter encompassed by the cancelled claims prior to this Amendment are not patentable over the art cited by the Examiner. Amendment and cancellation of these claims are made solely to facilitate expeditious prosecution of at least a portion of allowable subject matter in this application. We respectfully reserve the right to pursue claims, including the subject matter encompassed by the cancelled claims, as present prior to this Amendment and additional claims in one or more continuing applications.

Rejections under 35 U.S.C. §112

Regarding the definite nature of our recited term "the modified fields", we are referring to fields within records which were previously modified by a data cleaning process, as disclosed in our Abstract, and our paras. 0033, 0041, 0082 - 0084, 0086, 0090, and 0093 (referring to paragraph numbers as published by the USPTO).

We believe that the present amendment will clarify any possible misinterpretation what we are referring to by this term. We respectfully ask for the Examiner's reconsideration, and if believed still insufficient, we request the Examiner's suggestion for use of claim language in view of our disclosure paragraphs.

Rejections under 35 U.S.C. §101

These claims have been cancelled without prejudice from further consideration in this application.

Rejections under 35 U.S.C. §103(a)

We are unclear regarding the reason for the change in position by the Examiner since the first Office Action. In the previous rejections over Burdick in view of Wocke, the Examiner stated that Burdick failed to teach certain steps, elements and limitations of our claims, and employed Wocke to teach those missing claim aspects. We responded with arguments and explanation regarding our invention and our understanding of Wocke, but without amending our claims.

However, even though we did not amend our claims, in the present reasons for rejection, the Examiner has withdrawn Wocke from the proposed combination under 35 U.S.C. §103(a), and has indicated that Burdick *does* teach our claim aspects which he previously held were untaught (our emphasis added):

<u>Previous Office Action mailed on 11/14/2007</u>	<u>Current Office Action mailed on 5/21/08</u>
<p>As for Claim 1, Burdick et al <u>teaches</u></p> <p>"...declaring said data feature as suspect responsive to said degree of correlation exceeding a threshold" (see paragraph [0035], [0053]), <u>but fails to explicitly recite,</u></p> <p>"...generating a set of cleaning attributes..",</p> <p>"...receiving a data feature identified by a data mining process for a subset..", and</p> <p>"...determining a degree of correlation of said data feature to the modified fields".</p>	<p>As for Claim 1, Burdick et al <u>teaches,</u></p> <p>"declaring said data feature as suspect responsive to said degree of correlation exceeding a threshold" (see paragraph [0035], [0053]),</p> <p>"generating a set of cleaning attributes for each cleaned data record in a complete set of cleaned data records, said cleaning attributes reflecting which fields of each record have been modified by a cleaning operation" (see Fig. 1; see paragraph [0038]; e.g., attribute of an entity), and</p> <p>"determining a degree of correlation of said data feature to the modified fields of said subset of cleaned data records according to said cleaning attributes" (see paragraph [0035]; e.g., determining if two records are duplicates involves performing a similarity test that qualifies the similarity of two records),</p> <p><u>but fails to explicitly recite,</u></p> <p>"...receiving a data feature identified by a data mining process for a subset..".</p>

We agreed with the Examiner's initial position regarding what Burdick fails to teach, and we respectfully disagree with the Examiner's most recent position regarding what Burdick teaches.

Declaring A Data Feature as "Suspect". We have claimed:

"declaring said data feature as suspect responsive to said degree of correlation exceeding a threshold"

By "suspect", we mean the a data feature (e.g. a cluster or other feature) may be inaccurate or incorrect because the data it identifies has been "highly modified" by a previous cleaning process on the data (our para. 0041).

We respectfully disagree that Burdick teaches declaring "highly modified data from a cleaning process as being suspect" at their paragraphs [0035], [0053], which are (our emphasis added):

[0035] In the clustering and matching steps, algorithms identify and remove duplicate or "garbage" records from the collection of records. Determining if two records are duplicates involves performing a similarity test that quantifies the similarity (i.e., a calculation of a similarity score) of two records. If the similarity score is greater than a certain threshold value, the records are considered duplicates.

...

[0053] An output evaluation module 305 of the pre-processing component 202 evaluates the output of the three other functional modules 302, 303, 304. If the output is determined to be satisfactory, the output is passed to the automated learning component 203 via an output module 306. If the output is determined to be unsatisfactory (i.e., based on pre-defined thresholds, application-specific metrics, etc.), the three other functional modules 302, 303, 304 may be run again with different parameters. The output evaluation module 305 also may provide suggestions on how to change the execution of the three other functional modules 302, 303, 304 to improve the quality of the output (i.e., a feedback loop)

In paragraph 0035, we believe Burdick is merely detecting duplicates in the *uncleaned* data in a step to "pre-process" the data *before* cleaning is to be performed (para. 0051). Thus, their process would be unable to determine or declare a post-cleaning data feature as "suspect" as a result of previous cleaning operations.

Please consider that their "other functional modules" 302, 303, and 304 are, specifically, "Single-source Module", "Information Gathering Module", and "Planning Module", respectively.

Their *Single-Source Module* only combines sets of data from multiple sources to make them appear to be from a single source of data (para. 0050). We do not believe this would be considered a type of "data cleaning" by those ordinarily skilled in the art.

Their *Information Gathering Module* generates information about the record collection for input into an automated learning component (para. 0051). We do not believe this would be considered a type of "data cleaning" by those ordinarily skilled in the art, either.

And, their *Planning Module* estimates the resources (e.g. memory space) need to cleanse the record collection and create an execution plan for the cleansing process. But, this module is not disclosed as actually cleaning any data, it merely estimates resources in order to perform cleaning at some later stage of their invention.

We respectfully submit that Burdick's steps #302 - #305 (para. 0051) are part of a *preprocessing component* #202 (para. 0048), which are not performed *after* cleaning the data, but instead are performed *before* cleaning the data in order to prepare to perform cleaning, hence, their term *pre*-processing. Thus, the data could not be declared as "suspect" *as a result of cleaning* because cleaning has not been performed yet.

Cleaning Flags with Field-Granularity. We have also claimed:

"generating a set of cleaning attributes for each cleaned data record in a complete set of cleaned data records, said cleaning attributes reflecting which fields of each record have been modified by a cleaning operation"

Please refer to our previous reply which provides a more detailed explanation that our *cleaning attributes* are created to show which individual *fields* within a record have been modified.

It was reasoned in the current Office Action that Burdick actual does teach this claim element and its limitations at Fig. 1 and paragraph [0038] as an "attribute of an entity". We respectfully disagree.

Regarding Burdick's Figure 1, the term "attribute of an entity" does not appear, nor does "cleaning attribute" or just "clean". In fact, as best we can tell, the only "attribute" mentioned in Figure 1 is the "violated attribute dependancies" (eleventh row). We ask the Examiner to reconsider whether or not Figure 1 is actually illustrating a database (e.g. rows being records, columns being fields), or whether it is a tabular illustration of the *kinds* of problems that can occur in data and their possible solutions. Burdick's paragraph 0003 states only that Figure 1 illustrates the different "factors" which may result in dirty data, but it is silent regarding that Figure 1 illustrates actual dirty data (e.g. actual database contents).

Correlation Between Data Feature and Modified Data Fields. We have claimed:

"determining a degree of correlation of said data feature to the modified fields of said subset of cleaned data records according to said cleaning attributes"

Without our cleaning attributes which indicate which *fields* in each record have been modified by previous cleaning operations, Burdick cannot possibly or logically teach any other operations which depend on those cleaning attributes. So, for at least the foregoing reasons, we respectfully disagree with the Examiner's position that Burdick teaches this step at their para. 0035:

[0035] In the clustering and matching steps, algorithms identify and **remove duplicate or "garbage" records** from the collection of records. Determining if two records are duplicates involves performing a **similarity test** that quantifies the similarity (i.e., a **calculation of a similarity score**) of two records. If the similarity score is greater than a certain threshold value, the records are considered duplicates.

We believe this is merely a duplication detection method they are disclosing, but they are not correlating that detected duplication to whether or not those fields were changed by previous data cleaning. To know this, one would need an indicator of whether not those duplicate records had been modified by the previous cleaning operation, i.e. our cleaning attributes.

But, they have no field cleaning attributes, and thus their algorithm will simply delete the "duplicates" as they have determined them to be.

Our process, however, will detect duplicate records in which fields were also modified during earlier cleansing. Instead of simply deleting the duplicate records, our invention marks the cluster of duplicates as being suspect so that they can be reviewed. If the duplication was actually *caused by the data cleaning*, then deleting the duplicates may not be appropriate. Burdick's invention would delete them without further review, possibly causing an error in data handling which would be undiscoverable.

Please imagine a scenario such as this: a group of client transaction records for a retail web site include three transaction records in which the name fields contain "Bob Smith" and the city fields contain "Baltimore". But, now consider that 2 of the records for Bob Smith include street fields and zip code fields, but 1 record has no street or zip values:

Hypothetical Unclean Data					
<u>Record#</u>	<u>First Name</u>	<u>Last Name</u>	<u>Street</u>	<u>City</u>	<u>ZIP</u>
14	Bob	Smith	100 N Charles St	Baltimore	21201
132	Bob	Smith	100 N Charles Street	Baltimore	21201
564	Bob	Smith		Baltimore	
903	Bob	Smith	100 North Charles	Baltimore	21201
1143	Bob	Smith	100 N Charles St	Baltimore	21201

Now, a data cleaning process might make the street representations uniform (e.g. use "N" for "north" and "st" for "street"), and might fill in the "missing" street and zip code values in record #564, as follows:

Hypothetical <u>Cleaned</u> Data					
<u>Record#</u>	<u>First Name</u>	<u>Last Name</u>	<u>Street</u>	<u>City</u>	<u>ZIP</u>
14	Bob	Smith	100 N Charles St	Baltimore	21201
132	Bob	Smith	100 N Charles St	Baltimore	21201
564	Bob	Smith	100 N Charles St	Baltimore	21201
903	Bob	Smith	100 N Charles St	Baltimore	21201
1143	Bob	Smith	100 N Charles St	Baltimore	21201

Here's a considerable difference in how we believe Burdick's process would work compared to a process according to our claims. Burdick would detect records 132, 564, 903 and 1143 as duplicates of record 14 during their preprocessing (if they pre-processed cleaned data, which they do not), and Burdick's process would delete all but one of them.

But, what if record #564 was not really "missing" the 100 N Charles St. address values, but instead, the transaction represented by record #564 was to a different Baltimore address for Mr. Smith, perhaps to his home instead of his office? What would Burdick's invention do then?

Our invention would handle it by first adding our cleaning attributes to the records which might look something like this (using "1" to indicate a field that was changed):

Hypothetical Cleaned <u>and Tagged</u> Data						
<u>Record#</u>	<u>First Name</u>	<u>Last Name</u>	<u>Street</u>	<u>City</u>	<u>ZIP</u>	<u>C-Atts</u>
14	Bob	Smith	100 N Charles St	Baltimore	21201	00000
132	Bob	Smith	100 N Charles St	Baltimore	21201	00100
564	Bob	Smith	100 N Charles St	Baltimore	21201	00101
903	Bob	Smith	100 N Charles St	Baltimore	21201	00100
1143	Bob	Smith	100 N Charles St	Baltimore	21201	00000

Now, when our process compares the data feature (e.g. the seemingly duplicated records), and correlates that data feature to the our cleaning attributes, it can be seen that many

of the fields were modified by the cleaning process. So, it is not definite whether or not these are really duplicates, and thus they are declared "suspect" to be reviewed further, not simply to be deleted as Burdick would have done.

For these reasons, we respectfully ask the Examiner to reconsider the rejections, his previous arguments, and our present remarks. We believe Burdick fails to teach all of our claimed steps, elements, and limitations as relied upon in the rationale for rejections, and thus a *prima facie* case of obviousness under 35 U.S.C. §103(a) has not been established.

Request for Reconsideration of the Claim as a Whole

We have noticed that the Examiner's rationale for the rejections presents our steps in a different order than they are presented in our claims. We believe that this possibly has led to improper rejection of our claims by considering the steps individually, and not as a whole process.

The Federal Circuit has indicated that the claims must be considered as a whole, beyond analysis of only the differences between the individual claim components and multiple references:

[Although *Graham v. John Deere Co.*, 383 U.S. at 17, 148 USPQ at 476, requires that certain factual inquiries, among them the differences between the prior art and the claimed invention, be conducted to support a determination of the issue of obviousness, the actual determination of the issue requires an evaluation in the light of the findings in those inquiries of the obviousness of the claimed invention as whole, not merely the differences between the claimed invention and the prior art. *Lear Siegler, Inc. v. Aeroquip Corp.*, 733 F.2d 881, 221 USPQ 1025, 1033 (Fed. Cir. 1984) (emphasis added). See also *Fromson v. Advance Offset Plate, Inc.*, 755 F.2d 1549, 225 USPQ 26, 31 (Fed. Cir. 1985)

And:

It is impermissible to use the claimed invention, as an instruction manual or "template" to piece together the teachings of the prior art so that the claimed invention is rendered obvious. This court has previously stated that "[o]ne cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention." *In re Fritch*, 972 F.2d 2160, 23 USPQ2d 1780, 1784 (Fed. Cir. 1992) (quoting *In re Fine*, 837 F.2d 1071, 1075, 5 USPQ2d 1596, 16 (Fed. Cir. 1988)). See also *Akzo N.V. v. United States Int'l Trade Comm'n*, 808 F.2d 1471, 1 USPQ2d 1241, 1246 (Fed. Cir. 1986), *cert. denied*, 483 U.S. 909 (1987).

We have amended our claims in a manner which specifically states that cleaning of the data has already been performed before the first step of our claimed process. We respectfully request reconsideration of the rejections in view of our claims as a whole.

Request for Explicit Determination of Ordinary Skill Level

The Court in *KSR Int'l v. Teleflex Inc., et al.*, (U.S. Supreme Court, April 30, 2007) ("KSR") reiterated the importance of "resolving" the ordinary skill level using objective analysis when applying 35 U.S.C. §103(a) in a rejection, as set earlier forth in *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 17 - 18 ("Graham"). The Court in *KSR* clearly stated the need for *explicit* analysis (our emphasis added):

"... To determine whether there was an apparent reason to combine the known elements in the way a patent claims, it will often be necessary to look to interrelated teachings of multiple patents; to the effects of demands known to the design community or present in the marketplace; and to the background knowledge possessed by a person having ordinary skill in the art. **To facilitate review, this analysis should be made explicit.** . . ."

The reasons for rejection under 35 U.S.C. §103(a) as set forth in the Office Action did not include an explicit determination of what was the ordinary skill level in the art at the time of our invention. We believe this explicit determination by objective analysis is a requirement of the Examiner, not the Applicant, because an applicant is not required by law or rule to determine or state this fact, and the Examiner is a fact finder required to resolve *Graham* inquiries ("Examination Guidelines for Determining Obviousness Under 35 U.S.C. 103 in View of the Supreme Court Decision in *KSR International Co. v. Teleflex Inc.*, Fed. Reg., Vol. 72, No. 195, October 10, 2007). The Court has suggested a number of criteria which can be used to determine the ordinary skill level (*Environmental Designs, Ltd. v. Union Oil*, 713 F.2d 693, 696, 218 USPQ 865, 868 (Fed. Cir. 1983); *Bausch & Lomb, Inc. v. Barnes-Hind/Hydrocurve, Inc.*, 796 F.2d 443, 449-450, 230 USPQ 416, 420 (Fed. Cir. 1986)).

We believe it would be inappropriate to presume that the cited art shows the level or ordinary skill in the art. For example, Messrs. Burdick, Szczerba and Visgitus are listed as inventors of 20 issued US patents, according to the USPTO's online issued patent database searching using the query "*n/burdick-douglas\$ OR in/szczerba-robert\$ OR in/visgitus-joseph\$*". There is no evidence of record that "ordinary" persons in the art were such highly recognized innovators as these gentlemen.

For these reasons, we respectfully submit that we believe the cited art is drawn from inventors of *extraordinary* skill in the art, and thus their teachings do not indicate what was ordinary skill at the time of our invention. We are formally requesting an explicit determination

of the ordinary skill level at the time of our invention, in accordance with the Court's directions in *KSR*, *Graham*, and the USPTO's Examination Guidelines for Determining Obviousness Under 35 U.S.C. 103 in View of the Supreme Court Decision in *KSR*.

Request for Indication of Allowable Subject Matter

We believe we have responded to all grounds of rejection and objection, but if the Examiner disagrees, we would appreciate the opportunity to supplement our reply.

We believe the present amendment places the claims in condition for allowance. If, for any reason, it is believed that the claims are not in a condition for allowance, we respectfully request constructive recommendations per MPEP 707.07(j) II which would place the claims in condition for allowance without need for further proceedings. We will respond promptly to any Examiner-initiated interviews or to consider any proposed examiner amendments.

Respectfully,

/ Robert Frantz /

Robert H. Frantz
U.S. Patent Agent, Reg. N^o 42,553
Tel: (405) 812-5613
Franklin Gray Patents, LLC